SELECTION OF TEST ITEMS FOR DESIGNATED GOAL ATTAINMENT

Peter M. Ellis, Department of Business Administration, Utah State University, Logan, UT, 84322-3510
1-435-797-2372, peter.ellis@usu.edu
Ann L. Ellis, Department of Teacher Education, Weber State University, Ogden, UT, 84408-1304
801- 626-7343, aellis4@weber.edu

ABSTRACT

This work shows how items can be selected from a test bank so that an examination which seeks attainment of a given outcome goal might be constructed. The model presented here develops a general model which is easily adapted to the needs of a particular user. It permits consideration of separate test sections and the required inclusion of a certain number of items from each section. It shows how selection of competing items might be accomplished. It also provides for the generation of parallel tests, so that the identical examination is not used on consecutive test administrations.

INTRODUCTION

This approach envisions the existence of a test bank of a set of items which have been pretested. During test development item response theory (IRT) is applied in order to predict the probability that an examinee with a specific ability level will correctly answer a given item. A large pool of potential test items is field tested and item responses are analyzed for such indices as difficulty, discrimination power, reliability and validity. Items not meeting desired values are eliminated. Tests are then constructed by selecting items from the remaining pool.

For any item there might exist differing percentages of correct responses between participants who receive differing overall test scores. There will be a cohort of high achievers, a cohort of average achievers and another cohort of low achievers. For example, the high achieving group might be the top quartile, the middle achievers would come from the second and third quartiles and the low achievers would then come from the lowest quartile. One simple parameter is the index of discrimination **D**. It is determined as

$$\mathbf{D} = p_u - p_l$$

where p_u is the proportion in the upper group who answer the item correctly and p_l is the proportion of the lowest group who answer the item correctly.

The Mathematical Model

Several variables and parameters are needed in the model. They are:
D_{ics} = average correct response percentage on item i from cohort c in section s of the test
Ns = number of items to include from section s
<u>Ns</u> = specified value of Ns
ns = number of items available in the test bank for section s
X_{is} = 1 if item i in section s is of the test bank is included, and = 0 if not
C = number of cohorts to include
S = number of test sections to have

m = number of items from a restricted set that may be permitted for inclusion in the test
S_{cs} = aggregate score from cohort c in section s

Because several objectives are possible, the general objective function of the formulation will be designated as f($\underline{X}$), where $\underline{X}$ is the vector of the X_{is} values. With that, the general mixed bivalent integer formulation of the item selection problem is:

Problem (1)
maximize or minimize Z = f($\underline{X}$) (Ia)

subject to:

$$\sum_{i=1}^{ns} X_{is} = NS \qquad \text{(Ib)}$$

$$\sum_{i=1}^{ns} D_{ics}X_{is} - S_{cs} = 0 \qquad \text{(Ic)}$$

$$N_s = \underline{N_s} \qquad \text{(Id)}$$

all N_s, $S_{cs} \geq 0$
all $X_{is} = \{0,1\}$

This is a mixed bivalent integer formulation where the N_s and S_{cs} are continuous variables and the X_{is} variables are bivalent.

Constraints (Ib) establish the number of items that are to be chosen from each of the S test sections. The constraints of (Ic) calculate the expected total score for each cohort c. The limits provided in (Id) dictate the number of items that are to come from each section c. Note that the total number of items included on the test is the sum of the $\underline{N_s}$ values.

Several testing goals will be portrayed here. One of these is the maximization of the spread in expected scores between the top cohort and the bottom cohort. This comes about by writing Ia as

$$\text{maximize } z = \sum_{s=1}^{S} S_{Cs} - \sum_{s=1}^{S} S_{1s} \qquad \text{(Ia1)}$$

S_{cs} is the expected aggregate score from the top cohort in section s and S_{1s} is the expected aggregate score from the lowest cohort in any test section s. Therefore, (Ia1) defines the difference in expected scores between the top group and the bottom group. This objective function has the intent of providing the test that best discriminates between the high achievers and the low achievers.

A second objective to present is that of providing a test that maximizes the expected aggregate score of the lowest cohort. Such a test keeps all students as competitive as possible and minimizes the negative aspects of overall poor performance. The objective function f($\underline{X}$) is

$$\text{maximize} \quad z = \sum_{s=1}^{S} S_{1s} \qquad \text{(Ia2)}$$

Another version that deals primarily with the lowest cohort is to make the test as difficult for them as possible. Such a version might be of interest in a circumstance where there is a need to eliminate some candidates. This objective function would be

$$\text{minimize} \quad z = \sum_{s=1}^{S} S_{1s} \qquad \text{(Ia3)}$$

A fourth version of a goal might be to make the test as difficult as possible. The objective function would be written as

$$\text{minimize} \quad z = \sum_{s=1}^{S} S_{Cs} \qquad \text{(Ia4)}$$

Here the questions are selected with the goal that even the top cohort has the lowest possible aggregate score.

Often test preparers would like to have outcome consistency over several test experiences. This can come about by selecting a desired average score for all test takers. Denote this desired average as G. For instance, G = 80 would indicate that test questions should be selected so that the overall average for all test takers is 80. The number of test items will be N = N1 + … + NS. Therefore, the desired aggregate score is GxNxC for all included cohorts. Since this goal may not be perfectly attained, it is desirable to be as close to the desired aggregate total score as possible. If the actual aggregate score is above the target value GxN, the excess amount is designated as GP. If the actual aggregate is below the target value the shortage amount is designated as GM. The amount of deviation from the goal is written as

$$\sum_{c=1}^{C} \sum_{s=1}^{S} S_{cs} - GP + GM = GxNxC \qquad \text{(Ie)}$$

This new equation must be added to the constraint set. Then, the objective function is

$$\text{minimize} \quad z = GP + GM \qquad \text{(Ia5)}$$

CONCLUSION

Test item selection for the purpose of pursuing one of several performance goals has been illustrated here. A general mixed bivalent integer programming model has been developed and several objective functions have been included with it. Interested readers may contact the authors for the complete article.